DOCUMENT RESUME

ED 187 736 TM 800 231

AUTHOR Bachelor, Barry; And Others
TITLE Response Rarity as a Criterion in the Measurement of Originality.
PUB DATE Apr 80
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Creativity; *Creativity Tests; Elementary Education; Higher Education; *Responses; Scores; *Scoring; Test Reliability; *Test Validity
IDENTIFIERS Guilford Creativity Test Battery; Response Rarity; Wallach and Kogan Creativity Test Battery

ABSTRACT

Verbal responses to several tests of originality were subjectively rated for originality. The ratings were compared with the statistical frequency of these responses in two samples of test takers, 150 elementary school children and 60 college students. The elementary school children were administered three measures from the Wallach and Kogan test battery: Uses, Instances, and Similarity. The children were encouraged to think of a response nobody else in their class would think of. The college students' tests were the Alternate Uses test of Christensen, Guilford, Merrifield, and Wilson and the Consequences test of Christensen, Merrifield, and Guilford. Responses were classified in four categories: unique, used twice, used three times, and used four or more times. Responses were also rated on a five-point scale of originality. Interjudge reliability of the ratings ranged from .62 to .92 for various samples; but the correlation between originality ratings and infrequency of usage ranged from .03 to .16. It is suggested that response frequency scores are not a valid substitute for originality scores for tests like these. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED187736

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Response Rarity as a Criterion in the Measurement of Originality

Barry Bachelor

SWRL Educational Research and Development

Patricia Bachelor

University of Southern California

Dennis Hocevar

University of Southern California

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Bachelor, Barry
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM800 231

Paper presented at the 1980 conference of the American Educational Research Association, Boston, Mass., April 1980.

Abstract

The use of statistical infrequency (i.e., rarity) as a criterion for determining the originality of a response is investigated in two samples, 150 elementary school children and 60 college students. Using a variety of originality measures (e.g., Alternate Uses, Similarities, etc.), response rarity is shown to account for only a small percentage of the variance in subjective ratings of response originality. Such results indicate that alternatives to using statistical rarity as a criterion in the measurement of originality need to be considered.

Response Rarity as a Criterion in the Measurement of Originality

Within the creativity literature, a standard method for scoring originality tasks is to base weightings on statistical rarity. In other words, responses that appear infrequently are given more points while responses that occur frequently are given fewer points (e.g., Wallach & Kogan, 1965). Implicit in this scoring procedure is the assumption that statistically rare responses are original (i.e., clever and novel) as opposed to inappropriate or bizarre. While prior research has demonstrated that the correlation between a subjectively determined originality score (i.e., responses rated for quality) and an objectively determined originality score (i.e., one rated on the basis of statistical rarity) is strong (Hocevar, 1979b; Milgram, Milgram, Rosenbloom, & Rabkin, 1978) this should not be taken as evidence for the validity of statistical rarity scoring procedures. As Hocevar (1979a), Clark and Mirels (1972), and Hocevar and Michael (in press) have demonstrated, all methods of deriving originality scores are highly correlated with the total number of responses by virtue of the additive nature of the scoring formula. In other words, individuals who give more responses will tend to have higher originality scores regardless of the criteria of originality. Even originality scores based on random numbers are highly correlated with both subjectively and objectively determined originality scores (Hocevar, 1979b). The present study was designed to avoid the aforementioned problem with the scoring formula, and to further investigate the validity of using statistical rarity as a criterion for originality.

Method

Subjects

One hundred fifty children were randomly selected. Sampling was stratified so that thirty children were chosen for each of five age levels, six through ten. The children were white and from a middle class area. Approximately equal numbers of males and females were included.

Sixty college students were randomly selected from an introductory educational psychology class. All students were either juniors or seniors. Sampling was stratified so that an equal number of males and females were obtained.

Materials

The elementary school children were administered three measures from the Wallach and Kogan (1965) test battery including Uses, Instances, and Similarities. In the Uses test, children generate uses for each of five common objects (shoe, penny, etc.). In the Instances test, children generate instances of attributes (e.g., something that hurts). And in the Similarities test, children identify similarities between two common objects (e.g., a sidewalk and a street). In line with the recommendations of Harrington (1977), and Hocevar (1979a) the children were encouraged to think of creative responses (those that nobody else in their class would think of). Only their most creative response was solicited.

The college students were administered the Alternative Uses test (Christensen, Guilford, Merrifield, & Wilson, 1960) and the Consequences test (Christensen, Merrifield, & Guilford, 1958). The Alternate Uses test consists of three sets of three common objects. Examinees are given a common use for each object and then asked to "list as many as six others for which the object or parts of the object could serve." A time limit of

four minutes is given for each item (i.e., each set of three objects). There is no direction that the uses be original.

In the Consequences test, each item consists of a new and unusual situation, and examinees are given two minutes to think of consequences to that situation. Specifically, examinees are asked to "write as many different consequences or possible results of the change that you can." Conceptually, the test is similar to Torrance's (1974) Just Suppose test.

Scoring

Since the primary purpose of this study was to look at the validity of unique responses, scores for responses rather than individuals were computed. An advantage of this procedure is that when two or more originality scores for an individual are derived, they are spuriously correlated due to their common association with that individual's total number of responses. Further, averaging response ratings is not a reliable solution to this problem (Hocevar, 1979b). A frequency distribution for all responses was tabulated for each sample. For both samples, response frequencies were reduced to four categories: once, twice, three times, and four or more times. Each response was also subjectively rated for originality on a scale of 1(low) to 5(high). Original responses were defined as unique and clever. Ten judges were used for the elementary school children sample.

Because of the large number of responses (2,309) in the college student sample, judges were asked to subjectively rate only subsamples. Initially, a set of four judges rated the Alternate Uses responses for 30 students. A different set of judges was then asked to rate the responses for the remaining 30 students. A similar procedure was followed using the Consequences test. In summary, four different sets of four judges were used to rate the responses for the college student sample.

Results and Discussion

Response frequency category means are shown in Table 1. These means were tested by one-way ANOVAs. As shown in Table 1, significant category differences were found for both elementary school children's responses and college students' responses. Two aspects of the results seem to support the validity of using statistical infrequency as a criterion for originality. First, the trend was for statistically rare responses to get higher originality ratings and second, these differences are highly significant. However, eta squared (Glass & Stanley, 1970, 150), the proportion of variation attributable to categories, was only .17 for the elementary school children's responses and .07 or less for the college students' responses. The Pearson correlations between raw frequency scores and rated originality were similarly low in magnitude indicating that frequency accounts for 8 percent of the variance in rated originality for elementary school children's responses and no more than 4 percent of the variance for college students' responses. Such findings suggest that there is little relationship between the rated originality of responses and their statistical infrequency. In other words, many rare responses are not rated as original and vice versa. This finding should be disturbing to many researchers in the area of creativity since statistical rarity is often used as a criterion of originality.

Since subjective ratings of originality are reliable, easily obtained and intuitively more reasonable, their relatively infrequent use in the literature is puzzling. Treffinger and Poggio (1972) note that an alternative to statistical frequency needs to be considered. Subjective ratings are such an alternative, and a few studies (Hammaker, Shafto, & Trabasso, 1975; Hocevar, 1979ab; Malgady, & Barcher, 1977) point to the efficacy of this strategy.

Selected References

Clark, P., & Mirels, H. Fluency as a pervasive element in the measurement of creativity. Journal of Educational Measurement, 1970, 7, 83-86.

Hammaker, M., Shafto, M., & Trabasso, T. Judging creativity: A method for assessing how and by what criteria it is done. Journal of Educational Psychology, 1975, 67, 478-483.

Hocevar, D. Fluency as a confounding factor in the measurement of originality. Journal of Educational Psychology, 1979a, 71, 191-196.

Hocevar, D. A comparison of statistical infrequency and subjective judgment in the assessment of originality. Journal of Personality Assessment, 1979b.

Hocevar, D., & Michael, W. Effects of scoring formula on the discriminant validity of tests of divergent thinking. Educational and Psychological Measurement, in press.

Malgady, R., & Barcher, P. Psychological scaling of essay creativity: Effects of productivity and novelty. Journal of Educational Psychology, 1977, 69, 512-518.

Milgram, R., & Milgram, N. Creative thinking and creative performance in Israeli students. Journal of Educational Psychology, 1976, 3, 255-259.

Treffinger, D., & Poggio, J. Needed research on the measurement of creativity. Journal of Creative Behavior, 1972, 6, 253-267.

Wallach, M. & Kogan, N. Modes of thinking in young children: A study of the creativity-intelligence distinction. New York: Holt, Rinehart & Winston, 1965.

Winer, B. Statistical principles in experimental design. New York: McGraw-Hill, 1971.

Table 1

Mean Originality Ratings by Response Frequency Categories

	1	2	3	4 or more	Interjudge Reliability	F-Value	eta^2
Elementary School Sample	2.98(470)	2.49(116)	2.30(56)	2.18(135)	.92	51.20**	.16
College Students							
Alternate Uses (Subsample 1)	2.89(286)	2.66(63)	2.44(35)	2.28(97)	.82	11.43**	.07
Alternate Uses (Subsample 2)	2.88(344)	2.69(77)	2.84(34)	2.65(91)	.62	5.07*	.03
Consequences (Subsample 1)	3.00(508)	2.64(66)	2.30(27)	2.19(28)	.74	10.88**	.05
Consequences (Subsample 2)	2.54(501)	2.23(78)	2.15(26)	2.05(48)	.56	8.68**	.04

*p< .01

**p< .001

Note. Cell sizes in parentheses.